

## Contents

- [1 Introduction/Purpose of this Page](#)
- [2 Action Plan](#)
- [3 DMIT](#)  
[Comments/Discussion/Questions](#)
- [4 Gridded Datasets That Are Ready-To-Go](#)
- [5 Topics for Further Discussion](#)
- [6 Selected Resources and Notes](#)

## Introduction/Purpose of this Page

This page describes an action plan for developing the Unified Access Framework (UAF), a GEO-IDE project that is currently underway. Thanks to Steve Hankin for generating the initial idea. This project is being coordinated by [NOAA's Data Management Integration Team \(DMIT\)](#). A [UAF Technical Team](#) has been established to develop and implement the project.

The goal is to build momentum in GEO-IDE through an initial project that: 1) engages data providers, users, and IT folks; 2) leverages stable, proven solutions (one meaning of "standards"); and 3) has a high probability of demonstrable successes.

The basic idea is to develop a prototype implementation of a cross-NOAA gridded dataset integration capability.

This effort was discussed and refined at a [DMIT workshop in Silver Spring, MD on August 13 and 14, 2009](#). Plans for evolving the Unified Access Framework were discussed at a [technical meeting in Seattle, WA on November 16 - 18, 2010](#). If you have questions or are interested in participating in this activity in joining the DMIT please contact Ken McDonald (DMIT Chair.)

## Action Plan

[Presentation](#) describing the current status of the project as of January 14, 2010.

**Step 1: (FY2010) Build an integrated and well-documented collection of data and metadata that includes as many gridded datasets as possible and provide a unified access framework for that collection.**

**1-1 (lead ? Steve Hankin) Identify as many netCDF-CF-DAP (gridded) datasets as we can find in NOAA -- satellites, models, climatologies, bathymetries, HF radar, etc., that are currently available in THREDDS catalogs (a type of XML registry).** Note: At the August meeting the DMIT group agreed initially to confine efforts to servers and data sets known to the DMIT membership. This was deemed to be a sufficient core of data to develop and test the concepts.

**1-1a** Use wiki and expand to personalized data "server" call (DMIT) - ONGOING. See table below of identified datasets.

**1-2 (lead ? Steve Hankin) Create a single THREDDS catalog that integrates these existing catalogs into a single access point and enables all services that are readily available -- notably DAP, WMS and WCS.**

**1-2a** Create the THREDDS [Catalog](#) - COMPLETED. This is a "straw man" intended to be collectively added to and modified.

**1-2a.1** PMEL to host the server for the THREDDS XML document and make

## GEO-IDE\_Project\_Action\_Plan\_for\_FY\_2010

editable via SVN - COMPLETED. Since the [\[catalog\]](#) is visible to the public we don't want to do our experimental edits on it directly. A [UAF Grid Test Site](#) has been established to make changes.

**1-2a.2** Start with single structure for the catalog: by NOAA Line Organization/provider - COMPLETED.

**1-2a.3** DMIT Members expand dataset Listings [populate and coordinate to avoid duplication] - ONGOING.

**1-2b** PMEL will modify Java in house crawler to create "clean" TDS catalog from which various clients can be configured. NOTE: This will be a stop-gap solution until step 1.3b1 is achieved.

**1-2b.1** In Stage 1, valid datasets will be based upon whether JAVA netCDF can open them as scientific datasets. Invalid datasets will not be included

**1-2b.2** PMEL will make crawler code available to NGDC for their use in populating database describing GEO-IDE thredds catalog

**1-2b.3** Stage 2 will identify datasets suitable for aggregation and create NcML for those aggregations -- manually adding to the GEO-IDE THREDDS catalog. PMEL will explore a quasi-automated procedure based on enhancements to it's THREDDS crawler

**1-2c** Migrate the THREDDS Catalog to operational web presence

**1-2c.1** Select HOST for THREDDS Data Server

**1-2c.2** Move Catalog to THREDDS Data Server

**1-3** (*lead ? Ted Habermann*) **Harvest and improve metadata for selected datasets.**

**1-3a** Explore & examine additional sources of metadata to harvest

**1-3a.1** Ted, Ken M. and Lewis will evaluate existing metadata repositories for harvesting (e.g. GCMD GOS, NMMR)

**1-3b** Harvest metadata from the THREDDS catalog into database [see diagram](#)

**1.3b.1** Crawl the catalog and harvest selected metadata

◇ Evaluate what Australians have done in this area

◇ Explore PMEL ?crawling? and ?automation? capabilities

◇ Develop mechanism to harvest and develop other structures? thematic (GCMD or something else)

◇ And leading to TBD data discovery interfaces

**1-3c** Use expertise in metadata (FGDC/ISO, NCML, WCS/WMS Capabilities) to analyze and ?Repair/Extend? the harvested metadata (in close coordination with the dataset owners)

**1-3c.1** Ted and Steve to look into translating the harvested material into metadata to allow additional functions/structures/etc.

**1-3c.2** Add information to database to allow uniquely tailored outputs

**1-3c.3** Write NcML (both manually and from the existing metadata) to ?repair? datasets to allow thematic representation and provide other metadata as appropriate

**1-3d** Create authoritative "clean" catalog (replacing the version supplied by PMEL as a stop-gap in 1-2b.

**1-3e** Register UAF.Grid (components & services) with GEOSS, IOOS, and elsewhere that seems appropriate.

**1-4** (*lead ? Julie Bosch*) **Develop and document the ?process? (decision tree) for identifying/triaging and then including datasets into the catalog.**

**1-4a** Provide notification to providers when we ?pull? or ?link? or ?repair? their dataset

**1-4b** Need to decide when a dataset gets incorporated and then how it gets ?repaired? or improved

**1-4c** Establish Quality Control protocols

**1-4c.1** What metadata exists for this dataset?

**1-4c.2** Is dataset available on a DAP server?

**1-4c.3** Is dataset available through a THREDDS catalog already? If not, then create?

## GEO-IDE\_Project\_Action\_Plan\_for\_FY\_2010

**1-4c.4** Test datasets for compliance with known applications (get list of apps and tests from Steve, Rich, others...)

**1-4c.5** Is dataset CF compliant or not?

**1-4c.6** Does it pass CDM grid checker?

**1-4c.7** Is dataset COARDS compliant?

**1-4d** Establish Classification of datasets:

**1-4d.1** Tag with GCMD keywords (consider NERC instance of GCMD for this step)

**1-4d.2** Add standard names

**1-4e** Need to document mechanism for capturing information about the results of these tests into ISO compliant metadata record and how to maintain this as part of an ongoing data quality and monitoring process

**1-5** (*lead ? Roy Mendelsohn*) **Document how clients can connect/utilize the GEO-IDE resources -- desktop analysis and visualization tools, GIS systems, Web clients, etc.**

**1-5a** ERRDAP ? Roy M.

**1-5b** LAS ? Steve H.

**1-5c** ESRI/Environmental Data Connector (and others...)

**1-5d** MatLab - Rich S.

**1-5e** Google Earth - Kim J./John U.

**1-5f** Others...

**1-6** (*lead ? Ken McDonald*) **Build a web presence that explains GEO-IDE and points to the THREDDS catalog and associated tools.**

- Think about "branding" GEO-IDE as a service/resource for other projects...

- E.g. "GEO-IDE is providing the IOOS catalog of gridded ocean datasets."

**1-6a** Need parallel effort to have ?corporate? information to enable general users to understand the project

**1-6a.1** What is available

**1-6a.2** What you can do with it

**1-6a.3** Links to projects that are using the resource

**1-6b** Enhance GEO-IDE wiki to enable/encourage NOAA data management community to utilize, share, and enhance the THREDDS catalog

**1-6b.1** Documentation of ?GEO-IDE? examples showing how good things (science) were accomplished

**1-6c** Need way for general user to ?view? the data holdings

**1-6c.1** Google gadgets

**1-6d** Document several example finished application/product utilizing the THREDDS catalog

### Step 2: (FY 2010) Augment the testbed.

**2-1** In the GEO-IDE Wiki, document the standards and best practices used to create the gridded dataset collection, so that other data providers can add data to the collection.

**2-2** Work with data providers to add new data sources.

**2-2a** Apply Registration/Gate-keeping criteria to triage data and metadata for datasets that can be added to the collection

**2-3** Explore and implement time-aggregation of datasets at servers.

### Step 3: (FY2011+) Expand incrementally into to other data structures and other integration frameworks.

**3-1** Expand the offerings to include additional data collections. This includes observation types such as individual time series, individual vertical profiles and individual trajectories (e.g. TAO, OceanSites, tide gauges, SAMOS, ...)

**3-2** Add IOOS SOS inputs and outputs (and gateway services between SOS and DAP).

**3-2a** How can we leverage THREDDS catalog to do something with SOS?

**3-2a.1** How can we get to SOS service metadata into a THREDDS catalog?

**3-2a.2** Lay groundwork in year 1 to ensure that this will be possible...

**3-3 Add DAP Sequence services (and gateway services). This approach contrasts with 3.2 in that entire collections of observations become single "data objects" versus 3.2 in which individual degenerate grid data objects are connected only through their THREDDS catalog.**

## DMIT Comments/Discussion/Questions

- **From Ted H:**

On grids - I think this is a great idea and very worthwhile, BUT it is easy to underestimate the scope of the metadata task associated with making a big collection of NOAA grids understandable and useable, particularly for non-experts. I have looked at several of the existing examples and, frankly, have been very disappointed. In some cases, even experts in the discipline have been unable to discern the source or lineage of available datasets. We need to make sure that the plan includes active (and likely on-going) partnerships with data providers to make sure the data are understandable and useful to a broad spectrum of users. There is also the interesting question of OGC capabilities / service metadata... The capabilities documents that are auto-generated by these servers tend to be very bare-bones. We can definitely do better.

- **From Jeff DLB:**

The minimalist nature of OGC service metadata is a feature, not a bug. The intent was to not allow so much metadata therein that it becomes a de facto replacement for more complete metadata. There are slots for 0 or more Metadata URLs in the OGC service metadata, and these can be used to point to FGDC, ISO 19115 or other metadata. Unfortunately, most people use 0 of those slots.

- **From Steve H:**

On today's telcon we talked about an "informal data call" from DMIT as one bit of preparation that might be done in preparation for the summer GEO-IDE meeting. Below is a straw man -- including a few example responses.

- **From Julie B:**

For the metadata column I would add databased or managed information. In a request like this, we shouldn't limiting the idea of metadata to documents or type of standards. If they have it in a standard form great, if not we can work with them to develop/extract it.

- **From Deirdre J:**

I think OST/MDL has done some work on gridded weather data. I'll see if I can find someone who can contribute to this discussion topic, if that would help.

- **From Steve H:**

There are vast troves of GHRSSST (and other satellite) data that are served through OPeNDAP, but are not aggregated. The technical issue is that the files are huge and are bz2-compressed. To extract a time series from a collection like this means de-compressing each file in its entirety -- too slow and too costly to offer as a service. The technology to address this is new, but presumably works -- the use of aggregated netCDF4 "chunked" files. I suggest that a piece of the year 1 DMIT effort be used to pay someone to test this strategy. Thereafter (assuming success) to help satellite providers to serve their data as XYT aggregated datasets. The level of usability of the data would be revolutionized by this.

## Gridded Datasets That Are Ready-To-Go

*Question: Please tell us what gridded datasets would you be willing to share openly as part of an initial NOAA-wide "GEO-IDE" integration effort:*

| Dataset Name  | Description  | Format?<br>(netCDF,<br>CRIB, etc.)   | URLs and protocols where it is already served<br>Web pages? WCS? Other p   |
|---|--|--|--|
| ICOADS --<br>International<br>Comprehensive<br>Ocean-Atmosphere<br>Data Set                                 | Historical climatology<br>(1800-2007) of surface marine<br>variables: SST, SLP, winds,<br>cloud cover, humidity, data<br>density statistics, ...   | netCDF using<br>CF<br>conventions  | OPeNDAP:<br><a href="http://ferret.pmel.noaa.gov/thredds/dodsC/data/PM">http://ferret.pmel.noaa.gov/thredds/dodsC/data/PM</a>  |
| NODC Group for<br>High Resolution SST<br>(GHRSSST) Level 4<br>Gridded Products                              | L4 Data  | CF-compliant<br>netCDF (v3),<br>following<br>GHRSSST Data<br>Specifications<br>(GDS) | DAP: <a href="http://data.nodc.noaa.gov/pendap/ghrsst/L4">http://data.nodc.noaa.gov/pendap/ghrsst/L4</a><br><a href="ftp://data.nodc.noaa.gov/pub/data.nodc/ghrsst/">ftp://data.nodc.noaa.gov/pub/data.nodc/ghrsst/</a> HT<br><a href="http://data.nodc.noaa.gov/ghrsst/">http://data.nodc.noaa.gov/ghrsst/</a>  |
| NODC Pathfinder V5<br>SST   | L3 Data  | HDF4-SDS   | DAP: <a href="http://data.nodc.noaa.gov/pendap/pathfinder">http://data.nodc.noaa.gov/pendap/pathfinder</a><br><a href="ftp://data.nodc.noaa.gov/pub/data.nodc/pathfinder/">ftp://data.nodc.noaa.gov/pub/data.nodc/pathfinder/</a><br><a href="http://data.nodc.noaa.gov/pathfinder/">http://data.nodc.noaa.gov/pathfinder/</a>   |
| NODC World Ocean<br>Atlas 2005 (WOA)  | Gridded, analyzed,<br>3-dimensional fields   | netCDF (v3)  | DAP: <a href="http://data.nodc.noaa.gov/pendap/woa/WOA">http://data.nodc.noaa.gov/pendap/woa/WOA</a><br><a href="ftp://data.nodc.noaa.gov/pub/data.nodc/woa/WOA">ftp://data.nodc.noaa.gov/pub/data.nodc/woa/WOA</a><br><a href="http://data.nodc.noaa.gov/woa/WOA05nc/">http://data.nodc.noaa.gov/woa/WOA05nc/</a><br><a href="http://www.cdc.noaa.gov/thredds/dodsC/">http://www.cdc.noaa.gov/thredds/dodsC/</a> (restricted)<br><a href="http://www.cdc.noaa.gov/data/gridded/help.h">FTP: http://www.cdc.noaa.gov/data/gridded/help.h</a> |
| NESDIS Gridded<br>SST Products -- ...   | Products show global sea<br>surface temperatures in degrees<br>C at 14-, 50-, and 100-km<br>resolutions. They are composite<br>gridded-image data derived<br>from 8-km resolution Global<br>SST Observations and<br>generated on a global scale.<br>SST is defined as the skin<br>temperature of the ocean<br>surface water. | netCDF using<br>CF<br>conventions  | THREDDS: <a href="http://www.ngdc.noaa.gov/thredds/cat">http://www.ngdc.noaa.gov/thredds/cat</a>   |
| NMFS/SWFSC/ERD<br>Gridded Datasets -- ...   | 100's of datasets being served<br>by<br>NMFS/SWFSC/Environmetnal<br>Research Division.   | netCDF   | THREDDS: <a href="http://oceanwatch.pfeg.noaa.gov/thred">http://oceanwatch.pfeg.noaa.gov/thred</a>   |
| OceanNOMADS -<br>NOAA Operational<br>Model Archive and<br>Distribution System<br>for Ocean Models --<br>... | Operational ocean models from<br>NOAA (Real-Time Ocean<br>Forecast System-Atlantic) and<br>US Navy (Navy Coastal Ocean<br>Model and Intra Americas Sea<br>Nowcast-Forecast System).  | netCDF,<br>Grib2,<br>OPeNDAP   | <a href="http://edac-dap2.northerngulfinstitute.org/ocean_no">http://edac-dap2.northerngulfinstitute.org/ocean_no</a>  |

## GEO-IDE\_Project\_Action\_Plan\_for\_FY\_2010

|  |   |        |   |
|--|---|--------|---|
|  | Thirty-day rolling cache for each model.  |        |   |
| OFS model output files                             | The Operational Forecast Systems (OFS) have been created by NOAA/NOS/Co-Ops to provide the maritime community with improved short-term predictions of physical characteristics, such as water levels, currents, water temperatures and salinities, in specific estuaries. | netCDF | <a href="http://opendap.co-ops.nos.noaa.gov/thredds/catalog">http://opendap.co-ops.nos.noaa.gov/thredds/catalog</a> |
| Ocean Surface Current Analyses - Real time (OSCAR) | This project is developing a processing system and data center to provide operational ocean surface velocity fields from satellite altimeter and vector wind data.  | netCDF | <a href="http://dapper.pmel.noaa.gov/dapper/oscar/">http://dapper.pmel.noaa.gov/dapper/oscar/</a>                   |

### Topics for Further Discussion

- What can we mine to find existing resources (gridded datasets, DAP implementations, etc.) prior to the August workshop? Can we identify the datasets that we would like to include in an initial effort at the August workshop? Once those have been identified can we identify the best way to harvest, leverage, or create useful metadata to enable search and discovery?
1. OPeNDAP maintained list of OPeNDAP-Accessible Datasets
  2. UCAR maintained THREDDS Catalog
  3. NASA's Global Change Master Director OPeNDAP Portal
  4. NASA's Global Change Master Directory THREDDS Portal
  5. HF Radar THREDDS Catalog provided by NDBC
  6. NOAA's Environmental Model Activity Inventory
  7. NOMADS
- What can be achieved with respect to other elements in the proposed action plan including OGC services and searchable catalog?
  - How can we engage the CIO community (and specifically the IT security community) across NOAA to close the book on the security concerns?
  - How can we leverage/connect to ongoing work in the IOOS DIF project (e.g. Regional Data Provider THREDDS implementations)?
  - Should non-NOAA individuals/organizations (e.g. Unidata, OPeNDAP, OGC) with relevant technical expertise be invited to participate in the workshop?
  - Expanding the THREDDS catalog to include other services (e.g. SOS).
  - Explore the Google public data search and categorization projects (Jeff DLB to look into this.)
  - How should responsibilities/functions/activities be assigned across GEO-IDE and related (e.g. IOOS DIF) projects?

## Selected Resources and Notes

- Open-Source Project for a Network Data Access Protocol (OPeNDAP)
- Thematic Realtime Environmental Distributed Data Services (THREDDS)
- The DAP 2.0 protocol is an "endorsed standard" developed by NASA Earth Science Data Systems Standards Process Group
- Detailed documentation of The Data Access Protocol "DAP 2.0"
- Security alert and recommended resolution regarding a specific implementation of DAP 2.0
- DAP 2.0 is also an IOOS DMAC "recommended standard" and the IOOS DMAC standards website references the NASA information listed above.
- The combination of DAP 2.0, netCDF, and CF conventions is also documented as an IOOS DMAC "recommended standard."
- NetCDF overview presentation developed by Russ Rew of Unidata.
- NcML tutorial developed by Unidata
- Memo from NESDIS CIO regarding Practices for Securing Open-Source Project for a Network Data Access Protocol Server Software on NESDIS Information Systems, v2
- A method for setting up TDS catalogs for the first time developed by Rich Signell.
- Overview and benefits of the GEO-IDE UAF Grid Project